Full wwPDB X-ray Structure Validation Report (i)

Jan 6, 2024 – 09:01 pm GMT

PDB ID	:	6FZV
Title	:	Crystal structure of the metalloproteinase enhancer PCPE-1 bound to the procollagen C propeptide trimer (short)
Authors	:	Pulido, D.; Hohenester, E.
Deposited on	:	2018-03-15
Resolution	:	2.70 Å(reported)

This is a Full wwPDB X-ray Structure Validation Report for a publicly released PDB entry.

We welcome your comments at *validation@mail.wwpdb.org*
A user guide is available at
https://www.wwpdb.org/validation/2017/XrayValidationReportHelp
with specific help available everywhere you see the (i) symbol.

The types of validation reports are described at
http://www.wwpdb.org/validation/2017/FAQs#types.

The following versions of software and data (see references (i)) were used in the production of this report:

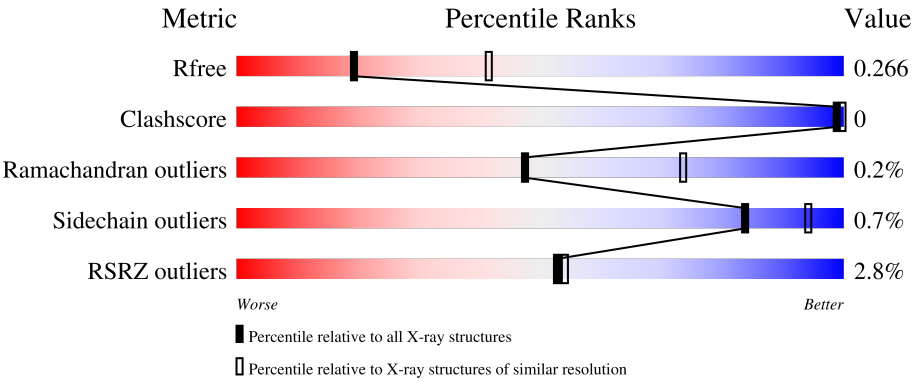
MolProbity	:	4.02b-467
Mogul	:	1.8.4, CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.36
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36

1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure:
X-RAY DIFFRACTION

The reported resolution of this entry is 2.70 Å.

Percentile scores (ranging between 0-100) for global validation metrics of the entry are shown in the following graphic. The table shows the number of entries on which the scores are based.



Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	2808 (2.70-2.70)
Clashscore	141614	3122 (2.70-2.70)
Ramachandran outliers	138981	3069 (2.70-2.70)
Sidechain outliers	138945	3069 (2.70-2.70)
RSRZ outliers	127900	2737 (2.70-2.70)

The table below summarises the geometric issues observed across the polymeric chains and their fit to the electron density. The red, orange, yellow and green segments of the lower bar indicate the fraction of residues that contain outliers for >=3, 2, 1 and 0 types of geometric quality criteria respectively. A grey segment represents the fraction of residues that are not modelled. The numeric value for each fraction is indicated below the corresponding segment, with a dot representing fractions <=5% The upper red bar (where present) indicates the fraction of residues that have poor fit to the electron density. The numeric value is given above the bar.

Mol	Chain	Length	Quality of chain
1	A	256	2% 94% 6%
1	B	256	3% 91% • 7%
1	C	256	% 90% • 9%
2	D	265	5% 87% • 11%

2 Entry composition (i)

There are 5 unique types of molecules in this entry. The entry contains 14456 atoms, of which 7049 are hydrogens and 0 are deuteriums.

In the tables below, the ZeroOcc column contains the number of atoms modelled with zero occupancy, the AltConf column contains the number of residues with at least one atom in alternate conformation and the Trace column contains the number of residues modelled with at most 2 atoms.

- Molecule 1 is a protein called Collagen alpha-1(III) chain.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	241	Total	C	H	N	O	S	0	0	0
			3711	1192	1820	323	363	13			
1	B	239	Total	C	H	N	O	S	0	0	0
			3671	1178	1799	320	361	13			
1	C	233	Total	C	H	N	O	S	0	0	0
			3585	1153	1755	313	351	13			

There are 39 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-10	GLU	-	expression tag	UNP P02461
A	-9	THR	-	expression tag	UNP P02461
A	-8	GLY	-	expression tag	UNP P02461
A	-7	HIS	-	expression tag	UNP P02461
A	-6	HIS	-	expression tag	UNP P02461
A	-5	HIS	-	expression tag	UNP P02461
A	-4	HIS	-	expression tag	UNP P02461
A	-3	HIS	-	expression tag	UNP P02461
A	-2	HIS	-	expression tag	UNP P02461
A	-1	SER	-	expression tag	UNP P02461
A	0	ALA	-	expression tag	UNP P02461
A	132	GLN	HIS	variant	UNP P02461
A	146	GLN	ASN	conflict	UNP P02461
B	-10	GLU	-	expression tag	UNP P02461
B	-9	THR	-	expression tag	UNP P02461
B	-8	GLY	-	expression tag	UNP P02461
B	-7	HIS	-	expression tag	UNP P02461
B	-6	HIS	-	expression tag	UNP P02461
B	-5	HIS	-	expression tag	UNP P02461
B	-4	HIS	-	expression tag	UNP P02461
B	-3	HIS	-	expression tag	UNP P02461
B	-2	HIS	-	expression tag	UNP P02461
B	-1	SER	-	expression tag	UNP P02461

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	0	ALA	-	expression tag	UNP P02461
B	132	GLN	HIS	variant	UNP P02461
B	146	GLN	ASN	conflict	UNP P02461
C	-10	GLU	-	expression tag	UNP P02461
C	-9	THR	-	expression tag	UNP P02461
C	-8	GLY	-	expression tag	UNP P02461
C	-7	HIS	-	expression tag	UNP P02461
C	-6	HIS	-	expression tag	UNP P02461
C	-5	HIS	-	expression tag	UNP P02461
C	-4	HIS	-	expression tag	UNP P02461
C	-3	HIS	-	expression tag	UNP P02461
C	-2	HIS	-	expression tag	UNP P02461
C	-1	SER	-	expression tag	UNP P02461
C	0	ALA	-	expression tag	UNP P02461
C	132	GLN	HIS	variant	UNP P02461
C	146	GLN	ASN	conflict	UNP P02461

- Molecule 2 is a protein called Procollagen C-endopeptidase enhancer 1.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
2	D	236	Total	C	H	N	O	S	0	0	0
			3428	1122	1660	293	344	9			

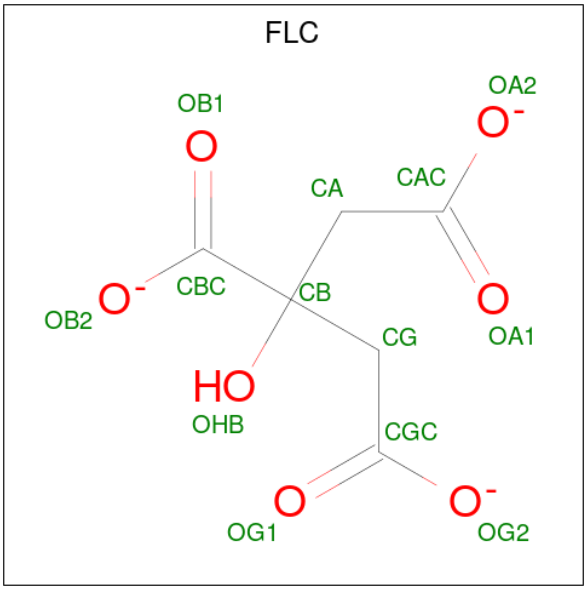
There are 12 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
D	-3	ALA	-	expression tag	UNP Q15113
D	-2	PRO	-	expression tag	UNP Q15113
D	-1	LEU	-	expression tag	UNP Q15113
D	0	ALA	-	expression tag	UNP Q15113
D	254	ALA	-	expression tag	UNP Q15113
D	255	ALA	-	expression tag	UNP Q15113
D	256	HIS	-	expression tag	UNP Q15113
D	257	HIS	-	expression tag	UNP Q15113
D	258	HIS	-	expression tag	UNP Q15113
D	259	HIS	-	expression tag	UNP Q15113
D	260	HIS	-	expression tag	UNP Q15113
D	261	HIS	-	expression tag	UNP Q15113

- Molecule 3 is CALCIUM ION (three-letter code: CA) (formula: Ca).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	A	1	Total Ca 1 1	0	0
3	B	1	Total Ca 1 1	0	0
3	C	1	Total Ca 1 1	0	0
3	D	2	Total Ca 2 2	0	0

- Molecule 4 is CITRATE ANION (three-letter code: FLC) (formula: $C_6H_5O_7$).



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	A	1	Total C H O 18 6 5 7	0	0
4	B	1	Total C H O 18 6 5 7	0	0
4	C	1	Total C H O 18 6 5 7	0	0

- Molecule 5 is CHLORIDE ION (three-letter code: CL) (formula: Cl).

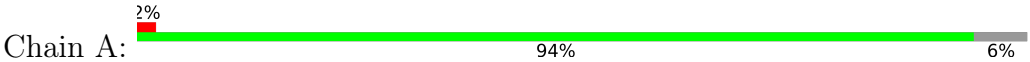
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
5	B	1	Total Cl 1 1	0	0
5	C	1	Total Cl 1 1	0	0

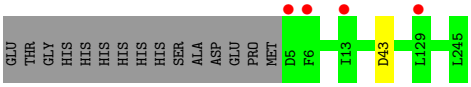
3 Residue-property plots (i)

These plots are drawn for all protein, RNA, DNA and oligosaccharide chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. Residues are color-coded according to the number of geometric quality criteria for which they contain at least one outlier: green = 0, yellow = 1, orange = 2 and red = 3 or more. A red dot above a residue indicates a poor fit to the electron density (RSRZ > 2). Stretches of 2 or more consecutive residues without any outlier are shown as a green connector. Residues present in the sample, but not in the model, are shown in grey.

• Molecule 1: Collagen alpha-1(III) chain

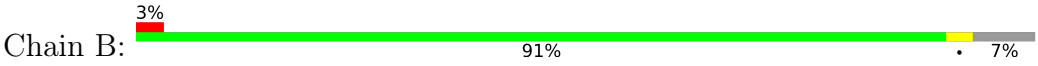
Chain A:

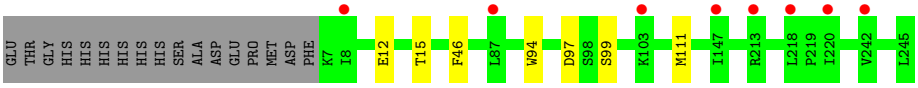




• Molecule 1: Collagen alpha-1(III) chain

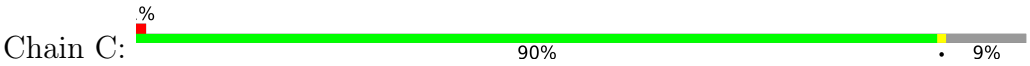
Chain B:

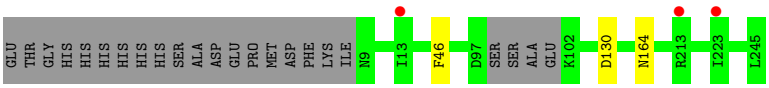




• Molecule 1: Collagen alpha-1(III) chain

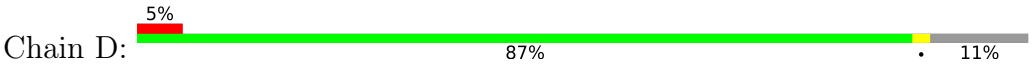
Chain C:

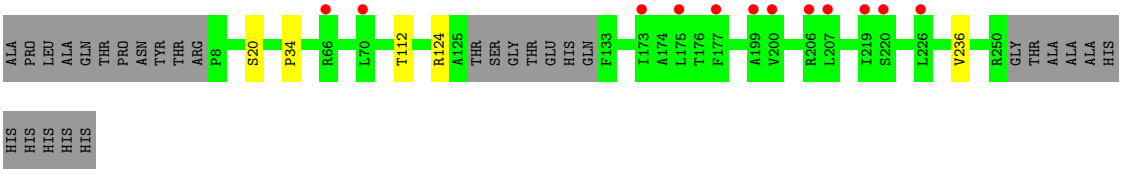




• Molecule 2: Procollagen C-endopeptidase enhancer 1

Chain D:





4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	89.09Å 144.12Å 158.74Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	79.37 – 2.70 106.70 – 2.70	Depositor EDS
% Data completeness (in resolution range)	99.8 (79.37-2.70) 99.8 (106.70-2.70)	Depositor EDS
R_{merge}	0.09	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.66 (at 2.69Å)	Xtriage
Refinement program	PHENIX 1.9_1692	Depositor
R, R_{free}	0.226 , 0.258 0.235 , 0.266	Depositor DCC
R_{free} test set	2827 reflections (4.98%)	wwPDB-VP
Wilson B-factor (Å^2)	82.8	Xtriage
Anisotropy	0.537	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.34 , 55.1	EDS
L-test for twinning[2]	$< \|L\| > = 0.49$, $< L^2 > = 0.32$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.94	EDS
Total number of atoms	14456	wwPDB-VP
Average B, all atoms (Å^2)	108.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.03% of the height of the origin peak. No significant pseudotranslation is detected.*

[1] Intensities estimated from amplitudes.

[2] Theoretical values of $< |L| >$, $< L^2 >$ for acentric reflections are 0.5, 0.333 respectively for untwinned datasets, and 0.375, 0.2 for perfectly twinned datasets.

5 Model quality

5.1 Standard geometry

Bond lengths and bond angles in the following residue types are not validated in this section: FLC, CA, CL

The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the expected value. A bond length (or angle) with $|Z| > 5$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Chain	Bond lengths		Bond angles	
		RMSZ	$\#\|Z\| > 5$	RMSZ	$\#\|Z\| > 5$
1	A	0.23	0/1936	0.40	0/2616
1	B	0.24	0/1916	0.40	0/2590
1	C	0.24	0/1873	0.40	0/2531
2	D	0.24	0/1816	0.42	0/2476
All	All	0.23	0/7541	0.40	0/10213

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no planarity outliers.

5.2 Too-close contacts

In the following table, the Non-H and H(model) columns list the number of non-hydrogen atoms and hydrogen atoms in the chain respectively. The H(added) column lists the number of hydrogen atoms added and optimized by MolProbity. The Clashes column lists the number of clashes within the asymmetric unit, whereas Symm-Clashes lists symmetry-related clashes.

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	A	1891	1820	1819	0	0
1	B	1872	1799	1799	3	0
1	C	1830	1755	1755	0	0
2	D	1768	1660	1660	1	0
3	A	1	0	0	0	0
3	B	1	0	0	0	0
3	C	1	0	0	0	0
3	D	2	0	0	0	0
4	A	13	5	5	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
4	B	13	5	5	1	0
4	C	13	5	5	0	0
5	B	1	0	0	0	0
5	C	1	0	0	0	0
All	All	7407	7049	7048	5	0

The all-atom clashscore is defined as the number of clashes found per 1000 atoms (including hydrogen atoms). The all-atom clashscore for this structure is 0.

All (5) close contacts within the same asymmetric unit are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:B:12:GLU:O	1:B:15:THR:OG1	2.10	0.67
4:B:302:FLC:OG2	4:B:302:FLC:OHB	2.18	0.58
1:B:97:ASP:OD2	1:B:99:SER:OG	2.28	0.50
1:B:94:TRP:HB3	1:B:111:MET:CE	2.47	0.45
2:D:20:SER:OG	2:D:124:ARG:NH2	2.52	0.42

There are no symmetry-related clashes.

5.3 Torsion angles (i)

5.3.1 Protein backbone (i)

In the following table, the Percentiles column shows the percent Ramachandran outliers of the chain as a percentile score with respect to all X-ray entries followed by that with respect to entries of similar resolution.

The Analysed column shows the number of residues for which the backbone conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles
1	A	239/256 (93%)	231 (97%)	8 (3%)	0	100 100
1	B	237/256 (93%)	226 (95%)	11 (5%)	0	100 100
1	C	229/256 (90%)	221 (96%)	8 (4%)	0	100 100
2	D	232/265 (88%)	211 (91%)	19 (8%)	2 (1%)	17 40
All	All	937/1033 (91%)	889 (95%)	46 (5%)	2 (0%)	47 73

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
2	D	112	THR
2	D	236	VAL

5.3.2 Protein sidechains (i)

In the following table, the Percentiles column shows the percent sidechain outliers of the chain as a percentile score with respect to all X-ray entries followed by that with respect to entries of similar resolution.

The Analysed column shows the number of residues for which the sidechain conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles
1	A	208/223 (93%)	207 (100%)	1 (0%)	88 96
1	B	206/223 (92%)	205 (100%)	1 (0%)	88 96
1	C	201/223 (90%)	198 (98%)	3 (2%)	65 86
2	D	189/217 (87%)	188 (100%)	1 (0%)	88 96
All	All	804/886 (91%)	798 (99%)	6 (1%)	84 94

All (6) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	43	ASP
1	B	46	PHE
1	C	46	PHE
1	C	130	ASP
1	C	164	ASN
2	D	34	PRO

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

5.3.3 RNA (i)

There are no RNA molecules in this entry.

5.4 Non-standard residues in protein, DNA, RNA chains (i)

There are no non-standard protein/DNA/RNA residues in this entry.

5.5 Carbohydrates (i)

There are no monosaccharides in this entry.

5.6 Ligand geometry (i)

Of 10 ligands modelled in this entry, 7 are monoatomic - leaving 3 for Mogul analysis.

In the following table, the Counts columns list the number of bonds (or angles) for which Mogul statistics could be retrieved, the number of bonds (or angles) that are observed in the model and the number of bonds (or angles) that are defined in the Chemical Component Dictionary. The Link column lists molecule types, if any, to which the group is linked. The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the expected value. A bond length (or angle) with $|Z| > 2$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	$\#\|Z\| > 2$	Counts	RMSZ	$\#\|Z\| > 2$
4	FLC	C	302	-	12,12,12	1.15	0	17,17,17	1.49	1 (5%)
4	FLC	A	302	-	12,12,12	1.16	0	17,17,17	1.52	1 (5%)
4	FLC	B	302	-	12,12,12	1.10	0	17,17,17	1.38	2 (11%)

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral centers analysed, the number of these observed in the model and the number defined in the Chemical Component Dictionary. Similar counts are reported in the Torsion and Rings columns. '-' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
4	FLC	C	302	-	-	8/16/16/16	-
4	FLC	A	302	-	-	7/16/16/16	-
4	FLC	B	302	-	-	12/16/16/16	-

There are no bond length outliers.

All (4) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
4	A	302	FLC	OB2-CBC-CB	4.30	120.52	113.05
4	C	302	FLC	OB2-CBC-CB	4.11	120.19	113.05
4	B	302	FLC	OB2-CBC-CB	3.61	119.32	113.05
4	B	302	FLC	OG2-CGC-CG	2.00	120.78	114.35

There are no chirality outliers.

All (27) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
4	A	302	FLC	CAC-CA-CB-CG
4	A	302	FLC	CAC-CA-CB-OHB
4	A	302	FLC	CG-CB-CBC-OB1
4	A	302	FLC	CG-CB-CBC-OB2
4	A	302	FLC	OHB-CB-CBC-OB1
4	A	302	FLC	OHB-CB-CBC-OB2
4	C	302	FLC	CAC-CA-CB-CG
4	C	302	FLC	CAC-CA-CB-OHB
4	C	302	FLC	CG-CB-CBC-OB1
4	C	302	FLC	CG-CB-CBC-OB2
4	C	302	FLC	OHB-CB-CBC-OB1
4	C	302	FLC	OHB-CB-CBC-OB2
4	A	302	FLC	CAC-CA-CB-CBC
4	C	302	FLC	CAC-CA-CB-CBC
4	B	302	FLC	CBC-CB-CG-CGC
4	B	302	FLC	OHB-CB-CG-CGC
4	B	302	FLC	CB-CG-CGC-OG2
4	B	302	FLC	CB-CG-CGC-OG1
4	B	302	FLC	CA-CB-CG-CGC
4	B	302	FLC	CA-CB-CBC-OB1
4	B	302	FLC	CA-CB-CBC-OB2
4	B	302	FLC	CG-CB-CBC-OB1
4	B	302	FLC	CG-CB-CBC-OB2
4	B	302	FLC	CAC-CA-CB-OHB
4	B	302	FLC	OHB-CB-CBC-OB1
4	B	302	FLC	OHB-CB-CBC-OB2
4	C	302	FLC	CA-CB-CBC-OB2

There are no ring outliers.

1 monomer is involved in 1 short contact:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
4	B	302	FLC	1	0

5.7 Other polymers (i)

There are no such residues in this entry.

5.8 Polymer linkage issues (i)

There are no chain breaks in this entry.

6 Fit of model and data

6.1 Protein, DNA and RNA chains

In the following table, the column labelled '#RSRZ> 2' contains the number (and percentage) of RSRZ outliers, followed by percent RSRZ outliers for the chain as percentile scores relative to all X-ray entries and entries of similar resolution. The OWAB column contains the minimum, median, 95^{th} percentile and maximum values of the occupancy-weighted average B-factor per residue. The column labelled 'Q< 0.9' lists the number of (and percentage) of residues with an average occupancy less than 0.9.

Mol	Chain	Analysed	<RSRZ>	#RSRZ>2			OWAB(Å^2)	Q<0.9
1	A	241/256 (94%)	0.51	4 (1%)	70	72	61, 88, 127, 205	0
1	B	239/256 (93%)	0.61	8 (3%)	46	46	60, 82, 132, 171	0
1	C	233/256 (91%)	0.50	3 (1%)	77	78	65, 92, 138, 170	0
2	D	236/265 (89%)	0.36	12 (5%)	28	26	74, 114, 184, 216	0
All	All	949/1033 (91%)	0.49	27 (2%)	53	54	60, 92, 161, 216	0

All (27) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	D	173	ILE	6.3
1	A	5	ASP	4.4
2	D	207	LEU	4.1
2	D	177	PHE	3.8
2	D	200	VAL	3.4
2	D	219	ILE	3.3
2	D	199	ALA	3.0
2	D	175	LEU	3.0
1	B	147	ILE	2.9
2	D	226	LEU	2.9
2	D	220	SER	2.9
1	A	13	ILE	2.8
1	B	8	ILE	2.6
1	B	242	VAL	2.6
2	D	66	ARG	2.4
1	C	13	ILE	2.4
1	B	213	ARG	2.3
1	A	6	PHE	2.3
1	A	129	LEU	2.3
2	D	206	ARG	2.2
1	B	87	LEU	2.1

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	B	218	LEU	2.1
1	B	103	LYS	2.1
1	C	223	ILE	2.1
1	C	213	ARG	2.1
2	D	70	LEU	2.0
1	B	220	ILE	2.0

6.2 Non-standard residues in protein, DNA, RNA chains (i)

There are no non-standard protein/DNA/RNA residues in this entry.

6.3 Carbohydrates (i)

There are no monosaccharides in this entry.

6.4 Ligands (i)

In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. The B-factors column lists the minimum, median, 95^{th} percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 0.9' lists the number of atoms with occupancy less than 0.9.

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å^2)	Q<0.9
4	FLC	B	302	13/13	0.81	0.20	89,94,114,114	0
4	FLC	C	302	13/13	0.84	0.15	88,96,115,115	0
4	FLC	A	302	13/13	0.85	0.13	84,95,117,117	0
5	CL	B	303	1/1	0.92	0.22	89,89,89,89	0
5	CL	C	303	1/1	0.94	0.08	99,99,99,99	0
3	CA	D	302	1/1	0.95	0.24	89,89,89,89	0
3	CA	C	301	1/1	0.95	0.27	80,80,80,80	0
3	CA	B	301	1/1	0.97	0.30	66,66,66,66	0
3	CA	A	301	1/1	0.97	0.28	60,60,60,60	0
3	CA	D	301	1/1	0.98	0.27	82,82,82,82	0

6.5 Other polymers (i)

There are no such residues in this entry.